# Moderationsfaktoren: ein Ansatz zur Analyse von Selektionsentscheidungen im Community Management

Paasch-Colberg, Sünje; Strippel, Christian; Laugwitz, Laura; Emmer, Martin; Trebbe, Joachim

Erstveröffentlichung / Primary Publication
Sammelwerksbeitrag / collection article

**Empfohlene Zitierung / Suggested Citation:**
Paasch-Colberg, S., Strippel, C., Laugwitz, L., Emmer, M., & Trebbe, J. (2020). Moderationsfaktoren: ein Ansatz zur Analyse von Selektionsentscheidungen im Community Management. In V. Gehrau, A. Waldherr, & A. Scholl (Hrsg.), *Integration durch Kommunikation (in einer digitalen Gesellschaft): Jahrbuch der Deutschen Gesellschaft für Publizistik- und Kommunikationswissenschaft 2019* (S. 109-119). Münster: Deutsche Gesellschaft für Publizistik- und Kommunikationswissenschaft e.V. https://doi.org/10.21241/ssoar.67858

DGPuK

Institut für Kommunikationswissenschaft
Westfälische Wilhelms-Universität Münster
DGPuK 2019 ifk

# Moderationsfaktoren: Ein Ansatz zur Analyse von Selektionsentscheidungen im Community Management

Sünje Paasch-Colberg, Christian Strippel, Laura Laugwitz, Martin Emmer & Joachim Trebbe

Institut für Publizistik- und Kommunikationswissenschaft, Freie Universität Berlin

**Zusammenfassung**

*Aggressive und diskriminierende Kommentare im Umfeld journalistischer Berichterstattung auf Nachrichtenseiten und in sozialen Medien gelten als Bedrohung für den gesellschaftlichen Zusammenhalt und Herausforderung für die verantwortlichen Redaktionen. Auf der Basis von 20 qualitativen Leitfadeninterviews mit Community Manager*innen untersucht dieser Beitrag, welche Moderationsstrategien im Umgang mit Hasskommentaren ergriffen werden und welche Faktoren diese Moderationsentscheidungen erklären können. Mit Rückgriff auf die Gatekeeping-Forschung werden die Ergebnisse zu einem Modell von Erklärungsfaktoren verdichtet, das diese ‚Moderationsfaktoren' auf den Ebenen des Individuums, der Profession, der Organisation und der Gesellschaft anordnet.*

**Keywords:** Community Management, Kommentarmoderation, Gatekeeping, Hate Speech, Interviews

 Moderationsfaktoren: Ein Ansatz zur Analyse von Selektionsentscheidungen im Community Management. In: Gehrau, V., Waldherr, A. & Scholl, A. (Hrsg.) Integration durch Kommunikation. Jahrbuch der Publizistik- und Kommunikationswissenschaft 2019, Münster, S.109-119.
DOI: https://doi.org/10.21241/ssoar.67858.

**Summary**

*Aggressive and discriminatory comments posted on news websites and social media threaten social cohesion and pose challenges for a site's respective editorial staff. Based on 20 qualitative, guided interviews with community managers, this article is an early examination of the factors that influence moderation decisions and the moderation strategies which are used to address hate comments online. By referencing preexisting gatekeeping research, this study models the explanatory factors which, in turn, define the ‚moderation factors' on the level of the individual, the norms and routines of the profession, the organization and society.*

**Keywords:** Community Management, Content Moderation, Gatekeeping, Hate Speech, Interviews

## Einleitung

Hetze, Hass und Diskriminierung haben die Sicht auf partizipative Medienangebote im Internet in den vergangenen Jahren deutlich verändert: Der anfänglichen Hoffnung auf eine Stärkung gesellschaftlicher Deliberation durch das Aufkommen von Kommentarspalten auf Nachrichtenseiten, Blogs und sozialen Netzwerken folgte der Befund, dass Diskussionen in diesen Kommunikationsräumen durch aggressive, beleidigende und diskriminierende Nutzerkommentare zunehmend gestört werden (Coe et al., 2014; Quandt, 2018; Su et al., 2018). Phänomene wie Cyber Mobbing, Inzivilität und Hate Speech haben dabei nicht nur negative Folgen für die von ihnen betroffenen Menschen, sondern bedrohen auch die gesellschaftliche Integration. So richtet sich etwa Hate Speech vor allem gegen marginalisierte Gruppen und deren Mitglieder und versucht sie von öffentlichen Diskussionen auszuschließen (Sirsch, 2013; Waltman & Mattheis, 2017). Da die Teilhabe und Repräsentation aller gesellschaftlichen Gruppen am bzw. im öffentlichen Diskurs als Voraussetzungen für kommunikative Integration gelten, stellt sich die Frage, wie in Redaktionen, die für die Gestaltung öffentlicher Debatten maßgeblich sind, mit solchen Hasskommentaren umgegangen wird.

In den Redaktionen hat sich mit dem vermehrten Aufkommen von problematischen Nutzerkommentaren eine neue journalistische Rolle professionalisiert: Community Manager*innen moderieren und selektieren Nutzerkommentare im Internet und regulieren so den öffentlichen Meinungsaustausch (Bakker, 2014; Paulussen, 2011; Braun & Gillespie, 2011). Diese Tätigkeiten werden in der Literatur oft als neue oder modifizierte Gatekeeping-Praxis beschrieben.

Die bisherige Forschung hat sich dabei vor allem den Arbeitsroutinen und Moderationspraktiken von Community Manager*innen (Reich, 2011; Binns, 2012; Chen & Pain, 2017; Frischlich et al., 2019) sowie den Folgen der Moderation insbesondere problematischer Kommentare (Binns, 2012; Ziegele & Jost, 2016) gewidmet. Die Befunde zu Moderationspraktiken im Umgang mit solchen Kommentaren sind bisher jedoch theoretisch unverbunden geblieben. Vor dem Hintergrund zunehmender Medienkritik und Zensurvorwürfen verlangt jedoch gerade dieser Aspekt der Kommentarmoderation eine systematische Betrachtung. Dieser Beitrag widmet sich diesem Desiderat und stellt ein theoretisches Modell zur Erklärung von Moderations- und Selektionsentscheidungen im Community Management vor. Um ein solches Modell erarbeiten zu können, stellen sich zwei Forschungsfragen:

FF1: Welche Strategien wenden Community Manager*innen in deutschen Redaktionen im Umgang mit Hasskommentaren an?

FF2: Anhand welcher Faktoren lassen sich die Moderations- und Selektionsentscheidungen erklären?

Im Rückgriff auf etablierte Gatekeeping-Modelle diskutiert dieser Beitrag Erklärungsfaktoren für Moderationsentscheidungen auf vier Ebenen: (1) jener des individuellen Community Managers, (2) professioneller Routinen, (3) der Organisation und (4) Gesellschaft. In Anlehnung an den etablierten Ansatz der Nachrichtenfaktoren (z. B. Eilders, 2006) und jüngere Arbeiten zu Diskussionsfaktoren (z. B. Ziegele, 2016) sprechen wir dabei von ‚Moderationsfaktoren'. Sie sollen erklären, wie nutzergenerierte Inhalte kategorisiert werden und welche journalistischen Selektions- und Moderationsentscheidungen daraus folgen. Zur Beantwortung der Forschungsfragen wurden Anfang 2018 qualitative Leitfadeninterviews mit Community Manager*innen deutscher Nachrichtenseiten im Internet geführt.

## Gatekeeping: Faktoren journalistischer Selektion

Ein Großteil der Forschung zum Community Management greift auf den in der Journalismusforschung etablierten Gatekeeping-Ansatz zurück. Dessen Ziel ist, den Prozess der journalistischen Informationskontrolle zu erklären (White, 1950). In der Literatur werden dazu verschiedene Formen des Gatekeepings diskutiert: So differenziert Rosengren (1974) etwa zwischen selektivem, qualitativem und quantitativem Gatekeeping – also der Auswahl, Aufbereitung und Gewichtung von Informationen zur Publikation.

Mit der Zeit wurden verschiedene Modelle zur Systematisierung der relevanten Einflussfaktoren auf journalistisches Verhalten vorgelegt, so etwa das Zwiebelmodell von Weischenberg (1992) oder die ‚hierarchy of influences' durch Shoemaker und Reese (1996). In einer synoptischen Gegenüberstellung verschiedener Modelle journalistischen Handelns und der jeweils berücksichtigten Erklärungsfaktoren macht Hanitzsch (2009) folgende sechs Analyseebenen aus: (1) Individuen, (2) Medienroutinen, (3) Organisationen, (4) Medienstrukturen, (5) Gesellschaft, (6) Kultur und Ideologie. Die meisten existierenden Modelle würden dabei vier oder fünf dieser Ebenen beinhalten (S. 156).


DOI: https://doi.org/10.21241/ssoar.67858.

Der Vergleich zeigt, dass alle Modelle eine *Individualebene* aufweisen, auf der Faktoren wie persönliche Merkmale und Einstellungen der Journalist*innen eingeordnet werden (Hanitzsch, 2009, S. 156-157). Auch die Ebene der *Medienroutinen* ist mit nur einer Ausnahme in allen Modellen angelegt. Hierunter fassen die jeweiligen Autor*innen neben professionellen Tätigkeitsmustern (wie etwa die Orientierung an Nachrichtenfaktoren) vor allem Faktoren wie Ressourcen und Zeitdruck (Hanitzsch, 2009, S. 156-157).

Abweichungen zwischen den verschiedenen Modellen bestünden hinsichtlich der Ebenen *Organisation* und *Medienstrukturen*, zwischen denen nicht alle Modelle unterscheiden (Hanitzsch, 2009, S. 156-157). Auf der Organisationsebene werden redaktionelle Strukturen, Abläufe und Ziele angesiedelt (Shoemaker & Reese, 2014, S. 9), während die Ebene der Medienstrukturen vor allem ökonomische Faktoren umfasse (Hanitzsch, 2009, S. 157). Gesellschaftliche, kulturelle und ideologische Einflüsse fasst Hanitzsch schließlich auf der Ebene der *Mediensysteme* zusammen. Diese Sphäre umfasst aus seiner Sicht medienpolitische und -rechtliche Faktoren sowie den nationalen kulturellen und sozialen Kontext (Hanitzsch, 2009, S. 157).

Wenngleich Journalismus, zumindest teilweise, auch heute noch ein massenmedialer Prozess der Informationsselektion und -verbreitung ist, so ist Nachrichtenproduktion im digitalen Zeitalter deutlich komplexer und kollaborativer: Die nicht-lineare Kommunikation mit Quellen und Publikum ist für den Prozess der Nachrichtenproduktion zentraler geworden. Zudem wurde ein Teil der Informationskontrolle von nichtjournalistischen Kommunikatoren, dem Publikum und technischen Gatekeepern (etwa digitalen Plattformen und Intermediären) übernommen, die nun ebenfalls Informationen produzieren, filtern, strukturieren und verteilen (Bro & Wallberg, 2015; Vos, 2015).

Trotz dieser dynamischen Verschiebung gilt der Gatekeeping-Ansatz noch immer als angemessenes Modell zur Analyse der Informationsproduktion im journalistischen Umfeld (Vos, 2015). Um den genannten Entwicklungen und den Leerstellen bestehender Gatekeeping-Modelle theoretisch gerecht zu werden, haben eine Reihe von Autor*innen modifizierte Modelle vorgeschlagen. So weist Bruns (2009) darauf hin, dass es dem Publikum in seiner hybriden Rolle als ‚produser' nun möglich ist, in großem Umfang Nachrichteninhalte zu kommentieren und zu ergänzen (S. 117). Die Tatsache, dass die meisten Webseiten ihre Kommentarspalten auf Post-Moderation umgestellt haben und die Community in die Moderation einbeziehen, deutet Singer (2014) als „another indication of a shift toward increased user ability to shape the content of news websites, with users making decisions about what others are to see or not see" (S. 60) und eine geteilte Gatekeeping-Verantwortung zwischen Journalist*innen und Nutzer*innen. Keyling (2017) ergänzt dies um den Hinweis, dass die aggregierten Aktivitäten der Nutzer*innen (z.B. Klicks, Likes, Kommentare) insbesondere Social-Media-Inhalte in einem kollaborativen Prozess gewichten, sichtbarer machen und damit die weitere Nutzung steuern (S. 79-81).

Barzilai-Nahon (2008) differenziert deutlicher als andere Autor*innen zwischen Gatekeepern und Gatekeeping-Mechanismen, die sie als „a tool, technology, or methodology to carry out the process of gatekeeping" definiert (S. 1496). Demnach ist die Kommentarmoderation also als neuer Gatekeeping-Mechanismus zu verstehen. Neben der Selektion von Kommentaren und Sprecher*innen können auch die verschiedenen Moderationspraktiken dazu gezählt werden, da sie Einfluss auf die Interaktion und Diskussionsqualität haben können und sich daher auch als eine Form der Inhaltsregulierung verstehen lassen. Eine Studie von Chen und Pain (2017) zeigt, dass sich dieser Gatekeeping-Mechanismus bereits normalisiert hat: Die Moderation wird von vielen Journalist*innen als neue berufliche Rolle verstanden, mit der versucht wird, die Informationskontrolle zurückzugewinnen, die durch das Auftreten der oben beschriebenen neuen Öffentlichkeitsakteure teilweise verloren ging.

## Community Management als neuer Gatekeeping-Mechanismus

Befragungsstudien weisen weitestgehend übereinstimmend auf ambivalente Einstellungen in den Redaktionen gegenüber nutzergenerierten Inhalten hin. Als besondere Herausforderungen für die neu entstandene Profession des Community Managements werden in der Literatur vor allem das hohe Aufkommen an Kommentaren, ihre teils mangelnde Qualität und fragwürdige Quellenlage sowie Bedenken hinsichtlich der Offenlegung von Quellen genannt (Diakopoulos & Naaman, 2011; Reich, 2011; Binns, 2012). Darüber hinaus zeigen diese Studien die Bandbreite der verschiedenen sozio-technologischen Lösungen, die zur Bewältigung dieser Herausforderungen eingesetzt werden, wie etwa ‚Netiquetten' für die Kommentierenden, unterschiedliche Registrierungsverfahren, automatische Filter, Pre- oder Post-Moderation sowie Bewertungssysteme und das technische Design der Seiten, die ange-


DOI: https://doi.org/10.21241/ssoar.67858.

messenes Verhalten belohnen oder Kommentare von positiv bewerteten Nutzer*innen priorisieren (Reich, 2011; Binns, 2012; Meltzer, 2015).

Die *Moderationspraktiken* im Umgang mit problematischen Kommentaren wurden grob in interaktive und nicht-interaktive Praktiken differenziert (Boberg et al., 2018). Als interaktive Praktiken gelten dabei etwa sachliche oder humorvolle Antworten auf problematische Kommentare und das Hervorheben wertvoller Benutzerkommentare. Hingegen sind das Löschen von Kommentaren und das Sperren von Nutzer*innen Beispiele für die nicht-interaktive Moderation. Eine dritte, kollaborative Strategie bezieht die Community der Nutzer*innen in die Moderation ein (Diakopoulos & Naaman, 2011; Ziegele & Jost, 2016).

Ein wachsender Fundus an Studien widmet sich der *Wirkung* solcher Moderationspraktiken auf die Nutzer*innen und die Diskussionsqualität (Binns, 2012; Ziegele & Jost, 2016; Kramp & Weichert, 2018; Ziegele et al., 2018). Sie weisen darauf hin, dass sich die wahrnehmbare Präsenz von Moderator*innen positiv auf die Nutzerdiskussionen auswirkt.

Erwähnenswert sind zudem zwei Studien, in denen tatsächliche Moderationsentscheidungen mithilfe einer automatisierten Inhaltsanalyse von gelöschten und publizierten Kommentaren auf Muster untersucht wurden. Die Befunde weisen auf Inkonsistenzen in der Moderationspraxis hin. Muddiman und Stroud (2017) analysieren über neun Millionen Nutzerkommentare der *New York Times* und können dabei zeigen, dass Kommentare mit bestimmten Schimpfwörtern signifikant häufiger gelöscht wurden als jene, die diese Worte nicht beinhalten. Ein ähnlich gerichteter, aber deutlich schwächerer Zusammenhang zeigte sich für Kommentare, die inzivile Begriffe aufweisen. Scheinbar lässt das Konzept der Inzivilität in der Moderationspraxis Spielraum für individuelle Interpretationen. In einer ähnlich konzipierten Analyse von Moderationsentscheidungen auf *Spiegel Online* (Boberg et al., 2018) ergab sich dagegen kein signifikanter Zusammenhang zwischen der Löschung eines Kommentars und der Präsenz von Schimpfwörtern. Was die Forscher*innen jedoch feststellen, ist ein Unterschied in der Moderation je Thema: Kommentare im Kontext der Themen Rechtspopulismus, Fake News und Geflüchtete wurden signifikant strenger moderiert.

Welche Faktoren letztlich zu welchen Moderationsentscheidungen führen, ist in der Literatur bisher nur vereinzelt untersucht und noch nicht in einem konsistenten Modell zusammengefasst worden. Neben den Kommentarinhalten werden dabei das Thema des Artikels (Reich, 2011; Kwon & Cho, 2017), Normen und Routinen (Muddiman & Stroud, 2017), individuelle Merkmale und das Rollenselbstverständnis als Moderator*in (Chen & Pain, 2017) sowie die redaktionelle Linie, das Image und wirtschaftliche Interessen (Binns, 2012; Pöyhtäri, 2014) als relevant angesehen. Im Folgenden greifen wir diese Thesen sowie die einzelnen Befunde auf und erarbeiten mit Rückgriff auf die bisherige Forschung zum Gatekeeping ein auch theoretisch fundiertes Erklärungsmodell für journalistische Entscheidungen in der Kommentarmoderation.

## Methode

Zur Beantwortung der Forschungsfragen wurden im Januar und Februar 2018 qualitative Experteninterviews mit Community Manager*innen von 20 deutschsprachigen Nachrichtenseiten und Social-Media-Angeboten durchgeführt. Die Auswahl der Befragten erfolgte dabei nach dem Prinzip des „maximum variation sampling“ (Schreier, 2018) entlang der Kriterien Reichweite, Angebotstyp, Finanzierung, Zielgruppe und Diskursarchitektur. Auf diese Weise sollte ein möglichst breites Spektrum an Erfahrungen mit Nutzerkommentaren abgedeckt werden. Berücksichtigt wurden die Internetseiten etablierter Printmedien mit nationaler Verbreitung, von Regional- und Lokalzeitungen, öffentlich-rechtlichen und privaten Radio- bzw. Fernsehprogrammen sowie originäre Internetangebote. Die Stichprobe bildet zudem verschiedene Diskursarchitekturen ab: 14 Angebote mit Kommentarspalten, drei mit Diskussionsforen und fünf Social-Media-Angebote. Zwei Angebote bieten sowohl Kommentarspalten als auch ein Diskussionsforum an und sind daher in Tabelle 1 zweifach ausgewiesen.

*Tabelle 1: Auswahlkriterien und Anzahl der realisierten Interviews (n=20)*

| Auswahlkriterium | Spezifikation | Interviews |
|---|---|---|
| **Medientyp** | Print | 13 |
| | Fernsehen | 3 |
| | Radio | 1 |
| | „Online only“ | 3 |
| **Finanzierung** | Öffentlich-rechtlich | 4 |
| | Privat | 16 |
| **Verbreitung** | Lokal/regional | 4 |
| | Überregional/national | 16 |
| **Diskussionsarchitektur** | Kommentarspalten | 14 |
| | Diskussionsforum | 3 |
| | Social Media | 5 |


DOI: https://doi.org/10.21241/ssoar.67858.

Die ausgewählten Redaktionen wurden kontaktiert und für ein Interview angefragt. In drei Fällen wurden auf Vorschlag bzw. Wunsch der Redaktion hin Doppelinterviews mit zwei Personen durchgeführt, sodass die Stichprobe 23 Personen umfasst. Auf dieser individuellen Ebene variieren die Eigenschaften der Interviewten in Bezug auf Geschlecht, Berufstätigkeit und Arbeitsposition. So wurden 13 Frauen und zehn Männer befragt. Aufgrund der unterschiedlichen Organisation und Größe der Redaktionen ist die Kommentarmoderation eine Aufgabe, die von Journalist*innen, Social-Media- oder Community-Manager*innen durchgeführt wird. Darüber hinaus finden sich in der Stichprobe auch Personen, die nicht selbst moderieren, aber Moderationsteam leiten. Neben der praktischen Erfahrung ließen sich so auch Informationen über Hintergründe der Moderation sammeln.

Der Interviewleitfaden umfasste 19 offene Haupt- und zehn offene Eventualfragen zu den Themenkomplexen (1) Arbeitsbedingungen und professionelles Selbstverständnis, (2) Kommentaraufkommen und Nutzerdiskussionen, (3) Hasskommentare sowie (4) Moderationsstrategien im Umgang mit Hasskommentaren. Die Interviews wurden transkribiert und in einer qualitativen Inhaltsanalyse mithilfe von MAXQDA ausgewertet. Dabei wurden deduktive und induktive Kategorienbildung kombiniert: In Anlehnung an das Modell von Shoemaker & Reese (1996; 2014), das international stark rezipiert wurde, haben die Analyseebenen Individuum, Profession und Routinen, Organisation und Gesellschaft die Codierung als deduktive Hauptkategorien angeleitet. Im Codierprozess wurden aus dem Material heraus Unterkategorien für die jeweiligen Erklärungsfaktoren entwickelt.

## Ergebnisse

Hasskommentare werden in allen von uns befragten Redaktionen als Herausforderung gesehen: Mehrere Interviewte gaben an, dass sie oder ihre Teams sich von der Menge und dem Ton der eingehenden Kommentare zeitweise überfordert fühlten. Hintergrund war dabei nicht selten ein Mangel an Personal. Entsprechend befanden sich auch einige der Teams zum Zeitpunkt der Gespräche in einer Umstrukturierung.

### *Moderationspraktiken und -strategien*

Um den genannten Herausforderungen zu begegnen, kommen in den Redaktionen eine Reihe verschiedener *Moderationspraktiken* zur Anwendung. Die Befragten berichten von nicht-interaktiven, kooperativen und interaktiven Maßnahmen und bestätigen damit den bisherigen Forschungsstand: Nicht-interaktive Praktiken umfassen das Bearbeiten und Löschen von Kommentaren, das Beobachten und Sperren von Nutzer*innen sowie das Schließen von Kommentarspalten (spontan oder systematisch für bestimmte Themen). Zudem arbeiten die meisten Redaktionen mit einer Software, die Kommentare mit bestimmten Begriffen automatisch verbirgt oder markiert. In einigen Fällen erlaubt die Software auch die Kooperation mit den Nutzer*innen, etwa indem sie problematische Kommentare melden können. Als interaktive Praktiken wurden das Verwarnen von Nutzer*innen, die sachliche oder humorvolle Gegenrede und das Loben wertvoller Kommentare genannt. Mit Blick auf die erste Forschungsfrage lassen sich anhand dieser Praktiken drei miteinander kombinierbare *Moderationsstrategien* identifizieren, die im Folgenen idealtypisch vorgestellt werden.

Zum ersten ließ sich bei manchen Redaktionen eine *Vermeidungsstrategie* beobachten, deren Ziel es ist, das Ausmaß insbesondere problematischer Kommentare stark einzudämmen. Dieses Ziel wird etwa durch einen beschränkten Zugang zur Diskussion (z. B. nur für Abonnent*innen einer Zeitung) oder die Deaktivierung der Kommentarfunktion erreicht. Auch lässt sich hierunter die Entscheidung fassen, die Kommentarfunktion nur bei weniger sensiblen Themen freizuschalten. Eine Interviewte berichtet davon, dass sensible Themen zu kritischen Tageszeiten nicht mehr auf die Webseite oder Social Media gestellt werden:

*„[W]ir sind uns den Themen ja bewusst und wissen, dass es gleich losgeht, wenn wir eines dieser Themen posten. Deswegen machen wir manche Themen auch gar nicht mehr auf Social [Media].“* (Nr. 14)

Die zweite von uns identifizierte Strategie fokussiert darauf, die Diskussion möglichst offen zu halten und *die Masse der Kommentare zu beherrschen*. Die ergriffenen Maßnahmen sind dabei oft nicht-interaktiv; zudem werden bestimmte Kommentare mithilfe von Tools automatisch geflaggt oder verborgen und Nutzer*innen beobachtet. Beispielhaft für diese Strategie steht etwa die folgende Beschreibung:

*„Wir sind eher unsichtbar tätig, weil wir eher löschend eingreifen oder Nutzer sperren. Direkt in den Dialog gehen wir eigentlich nur, wenn wir gezielt angesprochen werden, wenn es gezielt um uns geht, zum Beispiel wenn wir auf Fehler hingewiesen werden oder weil ein Kollege kritisiert wird, oder weil wir uns*


DOI: https://doi.org/10.21241/ssoar.67858.

*für einen Hinweis bedanken.“* (Nr. 1)

Mithilfe dieser Strategie können die Ressourcen auf die Moderation sensibler Themen konzentriert werden. Das Ziel ist dabei, möglichst früh zu intervenieren, Präsenz in der Diskussion zu zeigen und vergleichsweise streng zu moderieren.

Die dritte Strategie kann schließlich als *Bewältigungsstrategie* für die Moderator*innen bezeichnet werden. Sie wird zugleich auch als die hilfreichste Strategie zur Verbesserung des Diskussionsklimas beschrieben. Im Kern geht es dabei um den Versuch, die Ressourcen auf wertvolle Nutzerbeiträge zu konzentrieren und so Positives zu fördern. Eine Interviewte beschrieb die Vorteile dieser Strategie wie folgt:

*„Mittlerweile haben wir gemerkt, dass das gar nicht so gut tut, diesen destruktiven Postings so viel Aufmerksamkeit zu schenken, sondern eher die konstruktiven hervorzuheben. Jetzt geht es nicht nur uns besser, sondern gefühlt auch dem Forum und der Community. Gefühlt sind die Diskussionen konstruktiver und qualitätsvoller.“* (Nr. 18)

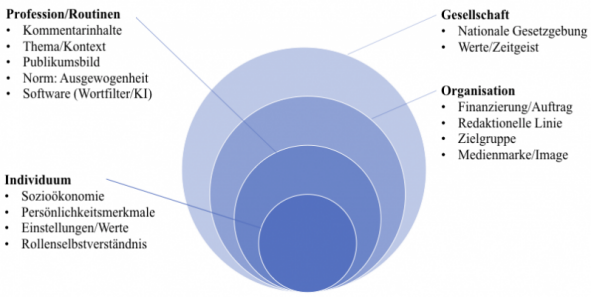
Redaktionelle Wertschätzung wird dabei auch durch die Seitengestaltung ausgedrückt, etwa durch das Hervorheben guter Kommentare als ‚Editor’s Pick‘ oder das Einbinden von Nutzerkommentaren in die Artikel. Andere Redaktionen berichten von erfolgreichen Sonderformaten wie etwa Live-Diskussionen per Video-Stream, mit denen direkte Begegnungen und das Diskussionsklima zwischen der Redaktion und den Nutzer*innen gefördert werden sollten. Diese Ergebnisse unterstützen den oben besprochenen Befund, dass die persönliche Präsenz der Moderation einen positiven Effekt auf die Diskussionskultur hat.

*Moderationsfaktoren*

Hinsichtlich der zweiten Forschungsfrage zeigen die Antworten der Befragten, dass die Moderations- und Selektionsentscheidungen in der Kommentarmoderation komplex sind. Auf den vier theoretischen Analyseebenen ließen sich jeweils mehrere relevante Erklärungsfaktoren ausmachen (siehe Abbildung 1).

*Individuelle Einflüsse* auf die Moderation werden von mehreren Befragten erwähnt: Während viele dieser Aussagen lediglich auf individuelle Unterschiede im Befinden gegenüber problematischen Kommentaren sowie im eigenen Moderationsstil verweisen, führen einige der Befragten explizit sozioökonomische Unterschiede und (Wert-)Einstellungen als die dahinter liegenden Faktoren an. Eine Befragte meinte etwa:

*Abbildung 1: Moderations- und Selektionsfaktoren für Nutzerkommentare*




*„Da wird es immer Unterschiede geben, je nachdem, was für einen Hintergrund man hat. Ist man eher ein junger Mensch, ein alter Mensch, ein Mann oder eine Frau. Da hat jeder ein anderes Empfinden.“* (Nr. 13)

Solche individuellen Unterschiede werden zwar kritisch betrachtet und es werden von einigen Redaktionen auch Maßnahmen ergriffen, um eine gemeinsame Linie in der Moderation herauszubilden; andererseits wird eine individuelle Moderation auch als authentisch wertgeschätzt, wie die folgenden Zitate zeigen:

*„Wir sind alle Menschen und man versucht natürlich, seine eigene Meinung ein bisschen zurückzunehmen. Das ist nicht immer einfach und wird nie ganz funktionieren.“* (Nr. 18)

*„[...] jeder, wie er kann. Ich meine, jemandem, der jetzt nicht besonders witzig ist, zu sagen: ‚Moderier das jetzt witzig ab‘, das würde ja eine Katastrophe werden. Und jemand, der nicht besonders sachlich ist, aber das Flapsige ganz gut kann, die sollen das so machen, wie sie es wollen und können.“* (Nr. 8)

Zudem zeigen die Interviews, dass sich *Routinen* in der Moderation herausbilden. Hinsichtlich der Kommentarinhalte werden etwa Beleidigungen, Schimpfworte, ein mangelnder Bezug zum Ausgangsartikel, Spam, Werbung und nicht-deutschsprachige Äußerungen nahezu übereinstimmend als Gründe für das Löschen von Kommentaren genannt. Dennoch ist die Bewertung der Kommentarinhalte kontextsensibel und auch vom Thema abhängig:

*„Wenn problematische Themen einlaufen, sind wir achtsam und geben dem auch die nötige Priorität. Pauschalisierende Kommentare und solche, die im*


DOI: https://doi.org/10.21241/ssoar.67858.

*Ansatz als hetzerisch wahrgenommen werden könnten, verschwinden dann sehr schnell. Bei solchen Themen moderieren wir definitiv strenger.“* (Nr. 6)

*„Wenn die Themen lockerer sind, lassen wir auch mal fünfe gerade sein.“* (Nr. 11)

Darüber hinaus ist das Objekt einer wertenden Aussage oft entscheidend für deren Einordnung. So zeigen sich einige Moderator*innen konsequenter, wenn eine gesellschaftlich marginalisierte Gruppe beleidigt wird, und weniger streng, wenn dies mit Journalist*innen geschieht. Zudem wurde berichtet, dass auch harmlosere Kommentare mal gelöscht werden, weil diese unter Umständen problematische Kommentare auslösen und zu einer Abwärtsspirale führen können. Solche Befunde zu individuellen Faktoren und kontextsensiblen Lösch-Routinen können die scheinbar inkonsistenten Befunde zur Moderationspraxis innerhalb einer Medienorganisation (z. B. Muddiman & Stroud, 2017; Boberg et al., 2018) erklären.

Des Weiteren umfasst die Ebene der Routinen auch das Publikumsbild, an dem sich die Moderationspraxis orientiert. Auf der einen Seite wurde der Wunsch geäußert, allen Nutzer*innen einen geschützten Raum anzubieten und daher strenger zu moderieren; andere Befragte betonen dagegen das Ideal der eigenständigen Meinungsbildung und eine dementsprechend weniger strenge Löschpraxis:

*„Damit lassen wir den Leuten die Wahl, mit wem sie mitgehen. Sie können den Ausgangspost sehen, sie können unsere Antwort sehen und jeder Leser darf sich selbst ein Bild machen und entscheiden, wie er damit umgeht. Ich glaube, die meisten Leser sind intelligent genug, die Entscheidung selbst zu treffen.“* (Nr. 13)

Daran anschließend steht die Norm der Ausgewogenheit, die manche Moderator*innen als Gefühl beschreiben, dass alle Meinungen ihre Berechtigung haben. Schließlich ordnen wir auf der Ebene der Routinen auch die Moderationssoftware und Wortfilter als eine Form von automatisierter Tätigkeitsroutine ein.

Auf Ebene der *Organisation* sind die Finanzierung eines Angebots und die damit verbundene gesellschaftliche Rolle relevante Moderationsfaktoren. Die Befragten aus den Redaktionen öffentlich-rechtlicher Angebote legten zum Beispiel offen, Kommentare und Nutzer*innen nur zögerlich zu löschen bzw. zu sperren und die Gründe zu dokumentieren. So antwortete eine Befragte aus dem öffentlich-rechtlichen Rundfunk:

*„Als öffentlich-rechtliche Anstalt ist Sperren richtig schwer. Die Leute zahlen halt Rundfunkbeiträge, da können wir eigentlich nicht so viel machen.“* (Nr. 8)

Im Gegensatz dazu verweisen private Angebote häufig auf ihr Hausrecht für die von ihnen verantworteten Internetseiten:

*„Grundsätzlich bin ich für das Hausrecht und ich finde, wir können veröffentlichen, was wir veröffentlichen wollen. [...] Ich halte den Eingriff für unser Recht. Das ist unsere Seite und das bestimmt auch, wie wir als Medium wahrgenommen werden.“* (Nr. 6)

Darüber hinaus ist auch die Redaktionslinie eines Angebots für die Moderationspraxis relevant. So wird etwa die politische Ausrichtung eines Nachrichtenangebots auf den Kommentarbereich übertragen:

*„Ein rechter Diskurs findet bei uns nicht statt, weil wir den gar nicht zulassen. Das ist bei anderen Zeitungen halt anders. Die Art, wie der Diskurs geführt oder wie stark ein rechter Diskurs zugelassen wird, spiegelt schon ein wenig die Haltung des Hauses wider, würde ich sagen.“* (Nr. 9)

Ein anderer Befragter berichtete, dass sich auch die Meinungsstärke eines Angebots in der Moderationspraxis widerspiegele, da etwa meinungsstarke Kommentare gefördert würden. Auffällig ist zudem, dass Bilder, GIFs und humorvolle Antworten im Rahmen der Moderation vor allem von Redaktionen eingesetzt werden, deren Angebote sich an ein jüngeres Publikum richten. Offensichtlich ist demnach auch die Zielgruppe ein relevanter Erklärungsfaktor für Moderationsentscheidungen.

Mit dem Strafrecht wurde schließlich auch ein Faktor der *gesellschaftlichen Makroebene* in den Interviews benannt. Alle Befragten geben an, strafrechtlich relevante Kommentare zu löschen und ggf. an verantwortliche Stellen weiterzugeben. Zu guter Letzt wird eine Art ‚Common Sense‘ der Gesprächskultur erwähnt, der in Form einer Netiquette festgehalten wird, aber an gesellschaftliche Vorstellungen davon anknüpfe, „wie man sich unterhalten möchte“ (Nr. 3).

Die Befragung zeigt damit insgesamt, dass Community Manager*innen auf eine Reihe relevanter Erklärungsfaktoren verweisen, wenn sie nach ihren Selektions- und Moderationsentscheidungen gefragt werden und dass sich diese Faktoren den in der Journalismusforschung etablierten Analyseebenen (1) Individu-


DOI: https://doi.org/10.21241/ssoar.67858.

um, (2) Profession und Routinen, (3) Organisation und (4) Gesellschaft zuordnen lassen. Entsprechend ergibt sich ein Modell von Moderationsfaktoren (siehe Abbildung 1), das an den theoretischen Forschungsstand anknüpft und in das sich neben den Ergebnissen dieser Studie auch die bisherigen Befunde zur Kommentarmoderation (siehe oben) einordnen lassen.

**Fazit**

Ziel dieser Studie war es, die Moderationsstrategien deutschsprachiger Redaktionen im Umgang mit diskriminierenden und hasserfüllten Nutzerkommentaren in Kommentarspalten, Diskussionsforen und Social Media aufzuzeigen und Faktoren zu benennen, die Unterschiede in der jeweiligen Moderationspraxis erklären können. Denn einerseits nimmt die neue Berufsrolle der Community Manager*innen eine gesellschaftlich relevante Gatekeeping-Funktion ein, wenn sie darüber entscheiden, ob ein Kommentar gelesen werden kann oder gelöscht wird. Andererseits steckt insbesondere die Forschung zu den Erklärungsfaktoren von Moderationsentscheidungen noch in den Kinderschuhen und entsprechende Befunde stehen bislang noch recht unverbunden nebeneinander.

Zur Schließung dieser Forschungslücke leistet diese Studie einen ersten Beitrag, in dem auf der Basis von Interviews mit Expert*innen ein Modell von Moderationsfaktoren entwickelt wird, das Erklärungsfaktoren auf der individuellen, professionellen, institutionellen und gesellschaftlichen Ebene umfasst. Dieses Modell kann dabei nur ein erster Vorschlag sein und bedarf weiterer empirischer Überprüfung und Ausarbeitung, auch um die methodischen Limitationen der Studie auszugleichen. Diese liegen – wie bei Befragungen üblich – in einer möglichen sozialen Erwünschtheit der Antworten sowie in dem Umstand, dass die Befragten nur solche Faktoren beschreiben können, derer sie sich bewusst sind. Nichtsdestotrotz kann das Modell als Ausgangspunkt dienen, um zukünftige Studien zur Kommentarmoderation systematischer als bisher zu verorten und theoretisch zu fundieren.

Darüber hinaus zeigen die Befunde, dass die Fragen nach angemessener Teilhabe und Repräsentation verschiedener gesellschaftlicher Gruppen, die sich im Umgang mit Hasskommentaren stellen, in den Redaktionen durchaus thematisiert werden. So besteht ein Spannungsverhältnis zwischen dem Ideal eines ‚safe space', in dem einer potentiellen Verdrängung marginalisierter Gruppen aktiv entgegengewirkt wird und dem Ideal eines Forums, in dem alle Ansichten öffentlich geäußert werden können. Darüber hinaus zeigt sich das Bewusstsein für potentiell (des-)integrierende Prozesse in Nutzerdiskussionen darin, dass die Moderator*innen teilweise strenger moderieren, wenn es um sensible Themen geht oder gesellschaftlich marginalisierte Gruppen in Kommentaren angegriffen werden. Zu guter Letzt öffnet die in den Interviews ausgedrückte Wertschätzung von Diversität innerhalb des Moderationsteams selbst Möglichkeiten für eine Diskussion darüber, wie Inklusion in den eigenen Reihen gelebt wird.

**Literatur**

Bakker, P. (2014). Mr. Gates returns. Curation, community management and other new roles for journalists. *Journalism Studies, 15*(5), 596–606. doi:10.1080/1461670X.2014.901783

Barzilai-Nahon, K. (2008). Toward a theory of network gatekeeping: A framework for exploring information control. *Journal of the American Society for Information Science and Technology, 59*(9), 1493–1512. doi:10.1002/asi.20857

Binns, A. (2012). Don't feed the trolls! Managing troublemakers in magazines' online communities. *Journalism Practice, 6*(4), 547–562. doi:10.1080/17512786.2011.648988

Boberg, S., Schatto-Eckrodt, T., Frischlich, L., & Quandt, T. (2018). The Moral Gatekeeper? Moderation and Deletion of User-Generated Content in a Leading News Forum. *Media and Communication, 6*(4). doi:10.17645/mac.v6i4.1493

Braun, J., & Gillespie, T. (2011). Hosting the public discourse, hosting the public. When online news and social media converge. *Journalism Practice, 5*(4), 383–398. doi:10.1080/17512786.2011.557560

Bro, P., & Wallberg, F. (2015). Gatekeeping in a Digital Era: Principles, practices and technological platforms. *Journalism Practice, 9*(1), 92–105. doi:10.1080/17512786.2014.928468

Bruns, A. (2009). Vom Gatekeeping zum Gatewatching. In C. Neuberger, C. Nuernbergk, & M. Rischke (Hrsg.), *Journalismus im Internet: Profession – Partizipation – Technisierung* (S. 107–128). Wiesbaden: VS.

Chen, G. M., & Pain, P. (2017). Normalizing Online Comments. *Journalism Practice, 11*(7), 876–892. doi:

10.1080/17512786.2016.1205954

Coe, K., Kenski, K., & Rains, S. A. (2014). Online and uncivil? Patterns and determinants of incivility in newspaper website comments. *Journal of Communication, 64*(4), 658–679. doi:10.1111/jcom.12104

Diakopoulos, N., & Naaman, M. (2011). *Towards quality discourse in online news comments.* Proceedings of the ACM 2011 conference on Computer supported cooperative work (S. 133–142). doi:10.1145/1958824.1958844

Eilders, C. (2006). News factors and news decisions: Theoretical and methodological advances in Germany. *Communications, 31*(1), 5–24. doi:10.1515/comm un.2006.002

Frischlich, L., Boberg, S., & Quandt, T. (2019). Comment Sections as Targets of Dark Participation? Journalists' Evaluation and Moderation of Deviant User Comments. *Journalism Studies, 20*(14), 2014–2033. doi: 10.1080/1461670X.2018.1556320

Hanitzsch, T. (2009). Zur Wahrnehmung von Einflüssen im Journalismus. Komparative Befunde aus 17 Ländern. *Medien & Kommunikationswissenschaft, 57* (2), 153–173. doi:10.5771/1615-634x-2009-2-153

Keyling, T. (2017). *Kollektives Gatekeeping: Die Herstellung von Publizität in Social Media*. Wiesbaden: VS.

Kramp, L., & Weichert, S. (2018). *Hass im Netz. Steuerungsstrategien für Redaktionen*. Leipzig: Landesanstalt für Medien NRW.

Kwon, K. H., & Cho, D. (2017). Swearing Effects on Citizen-to-Citizen Commenting Online. A Large-Scale Exploration of Political Versus Nonpolitical Online News Sites. *Social Science Computer Review, 35*(1), 84–102. doi:10.1177/0894439315602664

Meltzer, K. (2015). Journalistic Concern about Uncivil Political Talk in Digital News Media: Responsibility, Credibility, and Academic Influence. *The International Journal of Press/Politics, 20*(1), 85–107. doi:10.1177/1940161214558748

Muddiman, A., & Stroud, N. J. (2017). News Values, Cognitive Biases, and Partisan Incivility in Comment Sections. *Journal of Communication, 67*(4), 586–609. doi:10.1111/jcom.12312

Paulussen, S. (2011). Inside the newsroom: Journalists' motivations and organizational structures. In J. B. Singer, A. Hermida, D. Domingo, A. Heinonen, S. Paulussen, T. Quandt, Z. Reich, & M. Vujnovic (Hrsg.), *Participatory journalism: Guarding open gates at online newspapers* (S. 59–75). Oxford: Wiley-Blackwell.

Pöyhtäri, R. (2014). Limits of Hate Speech and Freedom of Speech on Moderated News Websites in Finland, Sweden, the Netherlands and the UK. *Annales, 24* (3), 513–524.

Quandt, T. (2018). Dark Participation. *Media and Communication, 6*(4), 36-48. doi:10.17645/mac.v6i4.1519

Reich, Z. (2011). User Comments: The transformation of participatory space. In J. B. Singer, A. Hermida, D. Domingo, A. Heinonen, S. Paulussen, T. Quandt, Z. Reich, & M. Vujnovic (Hrsg.), *Participatory journalism: guarding open gates at online newspapers* (S. 96–117). Oxford: Wiley-Blackwell.

Rosengren, K. E. (1974). International News: Methods, Data and Theory. *Journal of Peace Research, 11* (2), 145-156. doi:10.1177/002234337401100208

Schreier, M. (2018). Sampling and Generalization. In U. Flick (Hrsg.), *The SAGE Handbook of Qualitative Data Collection* (S. 84–97). London: SAGE.

Shoemaker, P., & Reese, S. (1996). *Mediating the message: Theories of influences on mass media content*. White Plains: Longman.

Shoemaker, P., & Reese, S. (2014). *Mediating the message in the 21st century: A media sociology perspective*. New York, London: Routledge.

Singer, J. B. (2014). User-generated visibility: Secondary gatekeeping in a shared media space. *New Media & Society, 16*(1), 55–73. doi:10.1177/1461444813477833

Sirsch, J. (2013). Die Regulierung von Hassrede in liberalen Demokratien. In J. Meibauer (Hrsg.), *Hassrede / Hate Speech: Interdisziplinäre Beiträge zu einer aktuellen Diskussion* (S. 165–193). Gießen: Gießener Elektronische Bibliothek.

Su, L. Y.-F., Xenos, M. A., Rose, K. M., Wirz, C., Scheufele, D. A., & Brossard, D. (2018). Uncivil and personal? Comparing patterns of incivility in com-

ments on the Facebook pages of news outlets. *New Media & Society, 20*(10), 3678–3699. doi:10.1177/1461444818757205

Vos, T. P. (2015). Revisiting Gatekeeping Theory during a Time of Transition. In T. P. Vos & F. Heinderyckx (Hrsg.), *Gatekeeping in Transition* (S. 3-24). London: Routledge. doi:10.4324/9781315849652

Waltman, M. S., & Mattheis, A. A. (2017). Understanding Hate Speech. *Oxford Research Encyclopedia of Communication*. doi:10.1093/acrefore/9780190228613.013.422

Weischenberg, S. (1992). *Journalistik. Theorie und Praxis aktueller Medienkommunikation. Band 1: Mediensysteme, Medienethik, Medieninstitutionen*. Opladen: VS.

White, D. M. (1950). The "Gate Keeper": A Case Study in the Selection of News. *Journalism Bulletin, 27*(4), 383–390. doi:10.1177/107769905002700403

Ziegele, M. (2016). *Nutzerkommentare als Anschlusskommunikation. Theorie und qualitative Analyse des Diskussionswerts von Online-Nachrichten*. Wiesbaden: Springer VS.

Ziegele, M., & Jost, P. B. (2016). Not Funny? The Effects of Factual Versus Sarcastic Journalistic Responses to Uncivil User Comments. *Communication Research*, (online first). doi:10.1177/0093650216671854

Ziegele, M., Jost, P., Bormann, M., & Heinbach, D. (2018). Journalistic counter-voices in comment sections: Patterns, determinants, and potential consequences of interactive moderation of uncivil user comments. *Studies in Communication and Media, 7*(4), 525–554. doi:10.5771/2192-4007-2018-4-525